

**Davies, JJ, Crain, RA, McCarthy, IG, Oppenheimer, BD, Schaye, J, Schaller, M and McAlpine, S**

**The gas fractions of dark matter haloes hosting simulated  $\sim L^*$  galaxies are governed by the feedback history of their black holes**

<http://researchonline.ljmu.ac.uk/id/eprint/10350/>

#### Article

**Citation** (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

**Davies, JJ, Crain, RA, McCarthy, IG, Oppenheimer, BD, Schaye, J, Schaller, M and McAlpine, S (2019) The gas fractions of dark matter haloes hosting simulated  $\sim L^*$  galaxies are governed by the feedback history of their black holes. *Monthly Notices of the Royal Astronomical Society*. 485 (3). pp. 3783-**

LJMU has developed [LJMU Research Online](http://researchonline.ljmu.ac.uk/) for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact [researchonline@ljmu.ac.uk](mailto:researchonline@ljmu.ac.uk)

# The gas fractions of dark matter haloes hosting simulated $\sim L^*$ galaxies are governed by the feedback history of their black holes

Jonathan J. Davies<sup>1</sup>,<sup>1</sup>★ Robert A. Crain<sup>1</sup>,<sup>1</sup> Ian G. McCarthy<sup>1</sup>,<sup>1</sup>  
Benjamin D. Oppenheimer<sup>2</sup>,<sup>2</sup> Joop Schaye<sup>3</sup>,<sup>3</sup> Matthieu Schaller<sup>3</sup>  
and Stuart McAlpine<sup>4</sup>,<sup>4</sup>

<sup>1</sup>*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*

<sup>2</sup>*CASA, Department of Astrophysical and Planetary Sciences, University of Colorado, 389 UCB, Boulder, CO 80309, USA*

<sup>3</sup>*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*

<sup>4</sup>*Department of Physics, University of Helsinki, Gustaf Hållstråmin katu 2a, 00560 Helsinki, Finland*

Accepted 2019 February 27. Received 2019 February 18; in original form 2018 October 15

## ABSTRACT

We examine the origin of scatter in the relationship between the gas fraction and mass of dark matter haloes hosting present-day  $\sim L^*$  central galaxies in the EAGLE simulations. The scatter is uncorrelated with the accretion rate of the central galaxy's black hole (BH), but correlates strongly and negatively with the BH's mass, implicating differences in the expulsion of gas by active galactic nucleus feedback, throughout the assembly of the halo, as the main cause of scatter. Haloes whose central galaxies host undermassive BHs also tend to retain a higher gas fraction, and exhibit elevated star formation rates (SFRs). Diversity in the mass of central BHs stems primarily from diversity in the dark matter halo binding energy, as these quantities are strongly and positively correlated at fixed halo mass, such that  $\sim L^*$  galaxies hosted by haloes that are more (less) tightly bound develop central BHs that are more (less) massive than is typical for their halo mass. Variations in the halo gas fraction at fixed halo mass are reflected in both the soft X-ray luminosity and thermal Sunyaev–Zel'dovich flux, suggesting that the prediction of a strong coupling between the properties of galaxies and their halo gas fractions can be tested with measurements of these diagnostics for galaxies with diverse SFRs but similar halo masses.

**Key words:** methods: numerical – galaxies: evolution – galaxies: formation – galaxies: haloes

## 1 INTRODUCTION

Analysis of the latest measurements of the primordial anisotropies exhibited by the cosmic microwave background (CMB) constrain the cosmic baryon fraction,  $\Omega_b/\Omega_0 \simeq 0.15$ , with sub-percent precision (Planck Collaboration 2018). Panoramic galaxy surveys indicate that approximately 5 per cent (by mass) of these baryons are in the form of stars and stellar remnants (see e.g. Balogh et al. 2001; Li & White 2009; Baldry et al. 2012), implying that the vast majority remains in gaseous form, leading to the notion of a ‘missing baryons’ problem.

In the absence of ejective feedback processes, one would expect the majority of the missing (i.e. non-stellar) baryons to be diffuse gas associated with collapsed haloes (e.g. Crain et al. 2007). However, feedback mechanisms that eject gas from the interstellar medium (ISM), and potentially also the haloes of galaxies, appear to be

a necessary component of galaxy formation models as a means to regulate galaxy growth (primarily via star formation feedback, e.g. Oppenheimer & Davé 2008; Crain et al. 2009; Oppenheimer et al. 2010; Schaye et al. 2010; Puchwein & Springel 2013), and to eventually quench their star formation and regulate the cooling of intragroup/intracluster gas via feedback associated with the growth of black holes (BHs) (e.g. Bower et al. 2006; Croton et al. 2006; Sijacki & Springel 2006; Fabjan et al. 2010; McCarthy et al. 2010). Estimates of the mass of circumgalactic gas based on absorption features in quasar sightlines or X-ray emission fall significantly short of the cosmic baryon fraction (e.g. Bregman 2007; Shull, Smith & Danforth 2012; Werk et al. 2013), and only in the massive haloes of rich galaxy clusters, with dynamical masses  $\sim 10^{15} M_\odot$ , is the baryon fraction observed to converge to the cosmic average (e.g. Allen, Schmidt & Fabian 2002; Lin, Mohr & Stanford 2004; Gonzalez et al. 2013).

Halo gas, often termed the circumgalactic medium (CGM), is the interface between galaxies and the intergalactic medium (IGM), and acts as reservoir of both the fuel for ongoing star formation, and

★ E-mail: j.j.davies@2016.ljmu.ac.uk

of the heavy elements whose synthesis accompanies this process. Numerical simulations of galaxy evolution demonstrate that the structure, temperature, element abundances, and ionization state of the CGM are markedly sensitive to the implementation of feedback processes (e.g. van de Voort & Schaye 2012; Hummels et al. 2013; Ford et al. 2016). Such processes remain poorly constrained by observations and their physical efficiencies cannot (yet) be predicted from first principles, meaning that *ab initio* prediction of the relationship between the gas fraction and total mass of haloes is not yet feasible.

Here we examine the influence of galaxy properties on the present-day gas fractions of haloes in the EAGLE simulations of galaxy formation. EAGLE adopts the pragmatic approach of calibrating feedback efficiencies to ensure the reproduction of key properties of the galaxy population, such as their stellar and central BH masses and the sizes of disc galaxies,<sup>1</sup> and has been shown to reproduce a diverse range of observable properties of the galaxy population (e.g. Furlong et al. 2015, 2017; Trayford et al. 2015, 2017; Segers et al. 2016; Crain et al. 2017) and intergalactic gas, as probed by X-ray emission (e.g. Schaye et al. 2015) and absorption features in quasar sightlines (e.g. Rahmati et al. 2015, 2016; Oppenheimer et al. 2016; Turner et al. 2017). The suite is therefore well suited to the study of the co-evolution of galaxies and their gaseous environments.

In Section 2, we briefly describe the simulations and our techniques for identifying and characterizing galaxies and their haloes. In Section 3, we present the scaling relation between the halo gas fraction and halo mass, and examine the origin of scatter about it, while in Section 4, we investigate means by which these predictions of the simulations can be confronted with observational measurements. We summarize and discuss our findings in Section 5. In Appendix A, we briefly compare the X-ray and thermal Sunyaev–Zel’dovich fluxes of our simulated haloes with current observational measurements. Throughout, we adopt the convention of prefixing units of length with ‘c’ and ‘p’ to denote, respectively, comoving and proper scales, e.g. cMpc for comoving megaparsecs.

## 2 METHODS

### 2.1 Numerical simulations

EAGLE (Evolution and Assembly of GaLaxies and their Environments; Crain et al. 2015; Schaye et al. 2015) is a suite of cosmological hydrodynamical simulations that follow the formation and evolution of galaxies.<sup>2</sup> EAGLE adopts a  $\Lambda$ CDM cosmogony described by the parameters advocated by the Planck Collaboration (2014), namely  $\Omega_0 = 0.307$ ,  $\Omega_b = 0.04825$ ,  $\Omega_\Lambda = 0.693$ ,  $\sigma_8 = 0.8288$ ,  $n_s = 0.9611$ ,  $h = 0.6777$ , and  $Y = 0.248$ . The simulations were evolved using a version of the smoothed particle hydrodynamics (SPH) and Tree-PM gravity solver GADGET-3 (last described by Springel 2005), incorporating several modifications to the hydrodynamics scheme. These include an implementation of the pressure-entropy SPH formulation of Hopkins (2013), the time-step limiter of Durier & Dalla Vecchia (2012), and switches for artificial viscosity and artificial conduction of the forms proposed by Cullen & Dehnen (2010) and Price (2010), respectively.

The EAGLE software also includes a number of subgrid treatments of processes operating below the numerical resolution of the

simulation, including radiative cooling (Wiersma, Schaye & Smith 2009a), star formation (Schaye & Dalla Vecchia 2008), stellar evolution and mass-loss (Wiersma et al. 2009b), BH formation, growth and mergers (Springel, Di Matteo & Hernquist 2005; Rosas-Guevara et al. 2015; Schaye et al. 2015) and feedback associated with the formation of stars (‘stellar feedback’, Dalla Vecchia & Schaye 2012), and the growth of BHs (‘AGN feedback’, Booth & Schaye 2009). The efficiency of stellar feedback was calibrated to reproduce the present-day stellar masses of galaxies and the sizes of galaxy discs, while the efficiency of AGN feedback was calibrated to reproduce the present-day scaling relation between the stellar mass of galaxies and that of their central BH (for further details see Crain et al. 2015). The gaseous properties of galaxies and their haloes were not considered during the calibration and may be considered predictions of the simulations.

We analyse four simulations from the EAGLE suite. We focus primarily on the simulation with the largest volume and greatest particle number, Ref-L100N1504, which evolves with the EAGLE Reference model a periodic cube of side  $L = 100$  cMpc, populated with  $N = 1504^3$  collisionless dark matter particles with mass  $9.70 \times 10^6 M_\odot$  and an (initially) equal number of baryonic particles with mass  $1.81 \times 10^6 M_\odot$ . In order to compute the intrinsic binding energy of haloes in this simulation, i.e. that which emerges in the absence of the dissipative physics of galaxy formation (see Section 2.2), we also analyse a simulation starting from identical initial conditions but considering only collisionless gravitational dynamics, DMONLY-L100N1504. We briefly examine NOAGN-L050N0752, a simulation following a smaller  $L = 50$  cMpc cubic volume at the same resolution, using a variation of the Reference model in which AGN feedback is disabled. To ensure that comparisons with this simulation are made on an equal footing we use Ref-L050N0752, a simulation of the same  $L = 50$  cMpc volume using the EAGLE Reference model. In all cases a Plummer-equivalent gravitational softening length of  $\epsilon_{\text{com}} = 2.66$  ckpc was used, limited to a maximum proper length of  $\epsilon_{\text{prop}} = 0.7$  pkpc.

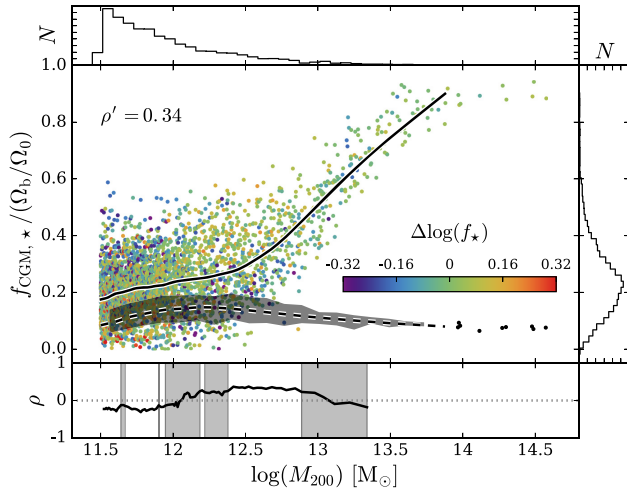
### 2.2 Characterizing haloes and galaxies

Haloes are identified by applying the friends-of-friends algorithm to the dark matter particle distribution, with a linking length of 0.2 times the mean interparticle separation. Gas, stars, and BHs are associated with the FoF group, if any, of their nearest dark matter particle. Bound substructures are subsequently identified within haloes using the SUBFIND algorithm (Springel et al. 2001; Dolag et al. 2009). We consider present-day haloes with  $M_{200} > 10^{11.5} M_\odot$ , with each halo thus being resolved by at least  $\sim 10^5$  particles. The typical present-day stellar mass of central galaxies hosted by haloes with  $M_{200} \simeq 10^{11.5} M_\odot$  is  $M_* \simeq 10^{9.5} M_\odot$ ; as shown by Schaye et al. (2015), present-day galaxies in EAGLE with at least this mass exhibit a passive fraction that broadly agrees with observational measurements.

We compute the spherical overdensity mass ( $M_{200}$ , Lacey & Cole 1994) of each halo about its most-bound particle, such that the mean density enclosed within the sphere of radius  $r_{200}$  is 200 times the critical density,  $\rho_c$ . More generally, halo properties are computed by aggregating the properties of all particles of the relevant type that reside within an appropriate aperture. We compute the inner halo binding energy by summing the binding energies of all particles within  $r_{2500}$  that comprise each halo’s counterpart in the DMONLY-L100N1504 simulation. In common with a number of other studies based on the analysis of EAGLE (e.g. Schaller et al. 2015; Matthee & Schaye 2019), we use this

<sup>1</sup>Section 2 of Schaye et al. (2015) motivates this approach in detail.

<sup>2</sup>The public release of EAGLE data is described by McAlpine et al. (2016).



**Figure 1.** Present-day halo gas fractions,  $f_{\text{CGM}}$ , as a function of halo mass,  $M_{200}$ . Histograms of  $M_{200}$  and  $f_{\text{CGM}}$  are shown above and to the right of the main panel, respectively. The solid curve denotes the running median, while the dashed curve denotes the running median of the total stellar fraction of the halo,  $f_*$ . The shaded region shows the 10th–90th percentile scatter of  $f_*$ . Symbols are coloured by the residuals of the relationship between  $f_*$  and  $M_{200}$ ,  $\Delta \log_{10} f_*$ . The lower panel shows the running value of the Spearman rank coefficient,  $\rho$ , for the  $\Delta f_{\text{CGM}} - \Delta \log_{10} f_*$  relation. The shading denotes where the recovered correlation is not significant ( $p > 0.01$ ).

‘intrinsic’ binding energy,  $E_{\text{DMO}}^{2500}$ , in order to eliminate the influence of dissipative baryonic processes. The dissipation of baryons in, and their ejection from, the progenitors of haloes throughout their formation and assembly can markedly influence their structure, potentially masking or exaggerating the influence of the intrinsic properties of the haloes. We pair haloes with their counterparts using the bijective particle matching algorithm described by Schaller et al. (2015), which successfully pairs 3411 of the 3543 haloes (96 per cent) satisfying  $M_{200} > 10^{11.5} M_{\odot}$  in Ref-L100N1504. Unpaired haloes are discarded, thus ensuring the same sample of haloes is used throughout.

Following Schaye et al. (2015), we compute the properties of central galaxies by aggregating the properties of the relevant particles that reside within 30 pkpc of the halo centre. We equate the BH mass of galaxies,  $M_{\text{BH}}$ , to the mass of their most-massive BH particle, which is almost exclusively coincident with the halo centre.

### 3 THE ORIGIN OF SCATTER IN HALO GAS FRACTIONS

In Fig. 1 we show, as a fraction of their mass  $M_{200}$ , the circumgalactic gas fraction,  $f_{\text{CGM}} \equiv M_{\text{gas}}(r < r_{200})/M_{200}$ , normalized by the cosmic baryon fraction, of present-day haloes in the Ref-L100N1504 simulation.<sup>3</sup> This definition includes the contribution of interstellar gas, but this is in general a small fraction of the halo gas mass. The solid black line shows the running median of the gas fraction,  $\tilde{f}_{\text{CGM}}(M_{200})$ , computed via the locally weighted scatterplot smoothing method (LOWESS; e.g. Cleveland 1979). There is considerable scatter in  $f_{\text{CGM}}$  in relatively low-mass haloes, which declines for  $M_{200} \gtrsim 10^{13} M_{\odot}$ .

<sup>3</sup>The baryon and stellar fractions of EAGLE haloes were presented and discussed by Schaller et al. (2015).

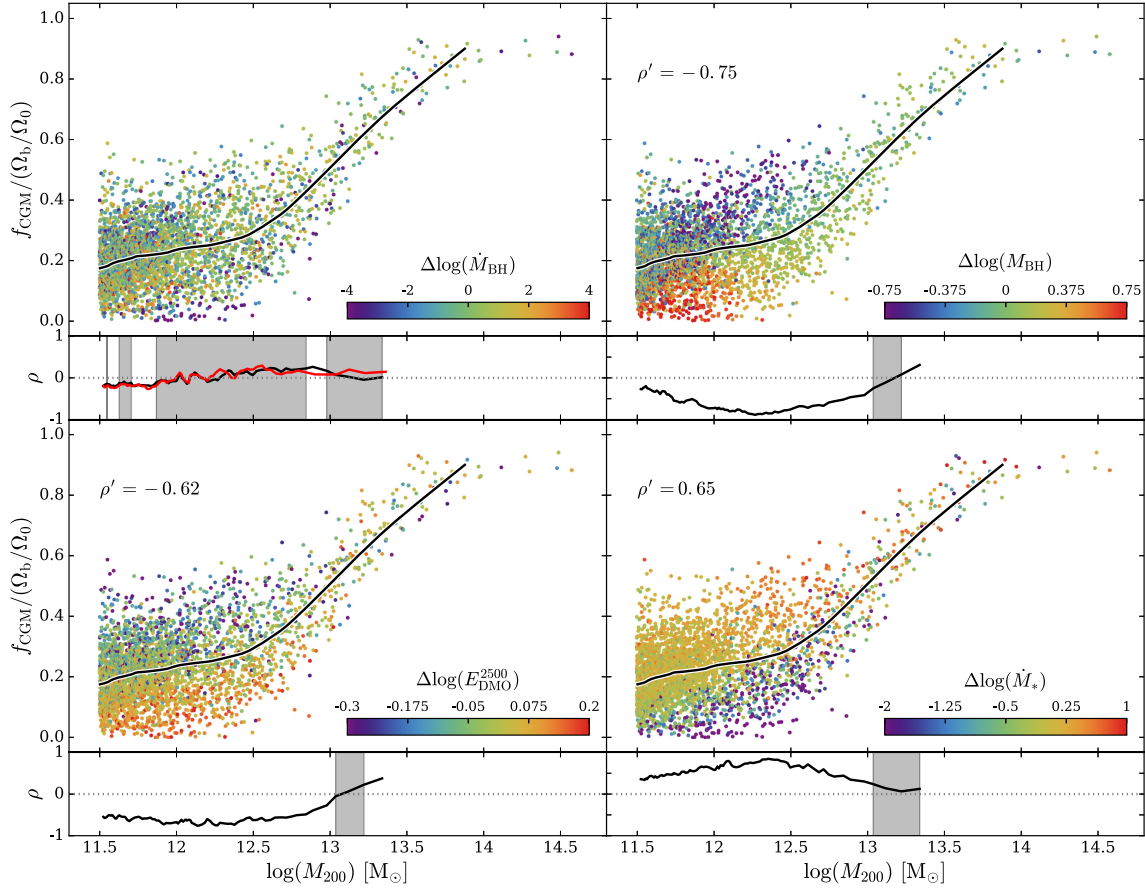
For reference, the running median of the total stellar-mass fraction of the halo,  $\tilde{f}_*(M_{200})$  is also shown as a dashed black curve, where  $f_* \equiv M_*(r < r_{200})/M_{200}$ . The shaded region about this curve denotes the 10th–90th percentile scatter of  $f_*$ . The LOWESS curves are plotted within the interval for which there are at least 10 measurements at both higher and lower  $M_{200}$ ; in poorly sampled high-mass bins, halo stellar fractions are plotted as individual black dots. Histograms of  $M_{200}$  and  $f_{\text{CGM}}$  are shown above and to the right, respectively, of the main panel. Gas fractions transition from  $\simeq 0.3 \Omega_b/\Omega_0$  below  $M_{200} \simeq 10^{12.5} M_{\odot}$ , rising steadily towards  $\simeq 0.9 \Omega_b/\Omega_0$  at  $M_{200} \simeq 10^{14} M_{\odot}$ , beyond which the trend flattens. This interval therefore represents a transition regime between EAGLE’s relatively gas-poor, low-mass haloes, and their gas-rich, high-mass counterparts. Present-day  $\sim L^*$  galaxies, with stellar mass similar to that of the Milky Way ( $M_* \simeq 6 \times 10^{10} M_{\odot}$ , e.g. Bland-Hawthorn & Gerhard 2016) are thought to be hosted by haloes with mass  $M_{200} \simeq 10^{12.5} M_{\odot}$  (e.g. Moster, Naab & White 2013).

Symbols are coloured by the residuals of the relationship between the stellar-mass fraction and the halo mass, i.e. for the  $i$ th halo,  $\Delta \log_{10} f_{*,i} = \log_{10} f_{*,i} - \log_{10} \tilde{f}_*(M_{200,i})$ . Haloes denoted by red (blue) points therefore have a greater (lower) stellar-mass fraction than is typical for their halo mass. Inspection of the symbol colours indicates that  $\Delta f_{\text{CGM}}$  and  $\Delta \log_{10} f_*$  are not strongly correlated at any mass scale. We quantify the strength of the correlation with the Spearman rank correlation coefficient,  $\rho$ . Since the correlations can in principle exhibit a strong dependence on halo mass, we compute ‘running’ correlation coefficients from halo-mass ordered sub-samples. For bins whose median halo mass exhibits  $M_{200} < 10^{12} M_{\odot}$ , we use samples of 300 haloes with starting ranks separated by 50 haloes (i.e. 1–300, 51–350, 101–400 etc), otherwise we obtain superior sampling of the high-mass range with bins of 100 haloes separated by 25 haloes.<sup>4</sup> This diagnostic is shown in the bottom panel of Fig. 1. In this and subsequent figures we shade regions where the  $p$ -value exceeds 0.01, corresponding to  $< 2.3\sigma$  confidence, to highlight where the recovered correlation is not significant. The correlation coefficient for the  $\Delta f_{\text{CGM}} - \Delta \log_{10} f_*$  is relatively low ( $|\rho| \lesssim 0.3$ ) at all halo masses, and is recovered with low confidence for much of the halo mass range. The coefficient for the 106 haloes in a 0.1 dex window about  $M_{200} = 10^{12.5} M_{\odot}$  can however be recovered with significance, and its value, which for reference we denote  $\rho'$ , is 0.34. This constitutes a weak-to-moderate correlation for a narrow range in  $M_{200}$ , but more broadly it is evident that the diversity of the gas fractions exhibited by present-day  $\sim L^*$  haloes does not emerge primarily as a consequence of some haloes converting more of their gas into stars than others. We note however that, as one approaches the ‘closed box’ regime of massive haloes with near-unity baryon fractions, one should expect a correlation between the scatter in  $f_{\text{CGM}}(M_{200})$  and  $f_*(M_{200})$  to emerge (Farahi et al. 2018).

The rest-mass energy liberated throughout the growth of central BHs is comparable to the gravitational binding energy of the halo (Silk & Rees 1998, see also Oppenheimer 2018), and is thus expected to foster the expulsion of gas from galaxies and their haloes. It is reasonable to surmise then that AGN feedback should have a significant influence on the gas fraction of haloes (e.g. Bower, McCarthy & Benson 2008; Puchwein, Sijacki & Springel 2008; McCarthy et al. 2010, 2011; Bocquet et al. 2016; Pillepich et al. 2018). In Fig. 2, we again show the present-day  $f_{\text{CGM}} - M_{200}$  relation of Ref-L100N1504, with the symbols in the top row coloured by

<sup>4</sup>The resulting coefficients are not strongly sensitive to these choices.





**Figure 2.** Present-day halo gas fractions,  $f_{\text{CGM}}$ , as a function of halo mass,  $M_{200}$ , with the solid curve denoting the running median. In each of the main panels, the symbols are coloured by residuals about the relationships of various properties as a function of halo mass: the accretion rate of the central BH ( $\dot{M}_{\text{BH}}$ , top left), the mass of the central BH ( $M_{\text{BH}}$ , top right), the intrinsic binding energy of the inner halo ( $E_{\text{DMO}}^{2500}$ , bottom-left), and the star formation rate of the central galaxy ( $\dot{M}_*$ , bottom-right). Beneath the main panels, we show the running value of the Spearman rank correlation coefficient for the relationships between these residuals and  $\Delta f_{\text{CGM}}$ , the residuals about the  $f_{\text{CGM}}-M_{200}$  relation. The red curve in the upper-left plot corresponds to the running  $\rho$  recovered when smoothing the BH accretion rate over a 100 Myr window. Shading denotes regions for which the correlation is recovered at low significance ( $p > 0.01$ ). In the cases where correlations are significant, we quote the Spearman coefficient,  $\rho'$ , of the correlation computed for haloes within a 0.1dex window about  $\log_{10} M_{200} [\text{M}_{\odot}] = 12.5$ .

the residuals about the median relation between the characteristics of central BHs and halo mass. In the top-left panel they are coloured by the residuals about the running median of the BH accretion rate as a function of halo mass,  $\Delta \log_{10} \dot{M}_{\text{BH}}$ . As in the case of the  $\Delta \log_{10} f_{\text{CGM}}$  correlation shown in Fig. 1, the Spearman rank correlation coefficient is relatively low ( $|\rho| < 0.3$ ) at all halo masses, and is recovered with low confidence for much of the sampled mass range, including at  $M_{200} = 10^{12.5} \text{M}_{\odot}$ , hence we are unable to quote a significant value of  $\rho'$  for this diagnostic. Scatter in halo gas fractions is therefore not strongly correlated with scatter in the BH accretion rate at any  $M_{200}$ . McAlpine et al. (2017) recently showed that the accretion rate of BHs in EAGLE can vary by orders of magnitude on very short time-scales ( $\lesssim 10^5$  yr), so we have repeated this test after time-averaging the BH accretion rate over the preceding 100 Myr. We find similar results with this definition of  $\dot{M}_{\text{BH}}$ , as is evident from its running value of  $\rho$ , shown as a red curve.

Prior analyses of the EAGLE simulations (Bower et al. 2017; McAlpine et al. 2018) have revealed that the development of a hot ( $T \gtrsim 10^6$  K), quasi-hydrostatic CGM in haloes with mass  $M_{200} \gtrsim 10^{12} \text{M}_{\odot}$  inhibits the buoyant transport away from the galaxy of gas ejected from the ISM in stellar feedback-driven outflows. The

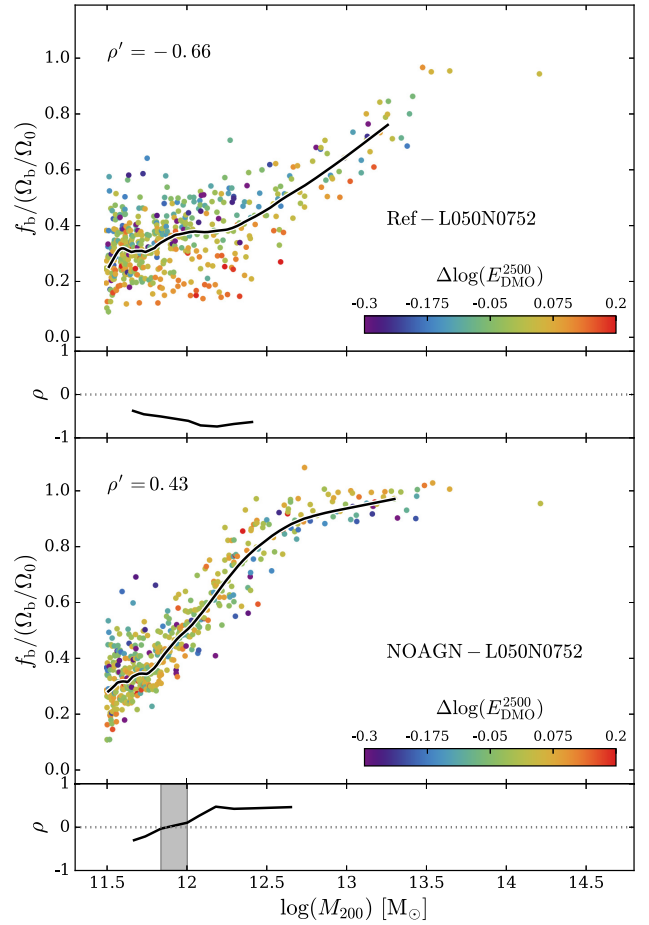
resulting build-up of gas triggers non-linear growth of the BH, which accretes rapidly until the feedback associated with its growth becomes the dominant means of regulating the inflow of gas on to the galaxy. McCarthy et al. (2011) argue that the expulsion of gas from the progenitors of group- and cluster-scale haloes, which accompanies this onset of BH feedback, occurs primarily at early cosmic epochs ( $1 \lesssim z \lesssim 3$ ) when their central BHs accreted most of their mass. We therefore colour the symbols of the top-right panel of Fig. 2 by the residuals about the running median of the BH mass as a function of halo mass,  $\Delta \log_{10} M_{\text{BH}}$ . In this case, the colouring reveals a striking negative correlation between the gas fraction (at fixed halo mass) and the BH mass, such that haloes whose central galaxies host atypically massive central BHs exhibit systematically low gas fractions, and vice versa. The visual impression is corroborated by the Spearman rank correlation coefficient, which is significant and negative for all  $10^{11.5} < M_{200} \lesssim 10^{13} \text{M}_{\odot}$ . The coefficient for haloes with  $M_{200} \simeq 10^{12.5} \text{M}_{\odot}$  is very strong,  $\rho' = -0.75$ . These results indicate that the halo gas fractions of  $\sim L^*$  galaxies are regulated primarily by the evolutionary stage of their central BHs.

We turn next to the origin of the diversity of BH masses at fixed  $M_{200}$ . This question was explored by Booth & Schaye (2010, 2011) using the OWLS simulations (Schaye et al. 2010), who concluded

that BH mass is governed primarily by the binding energy of the inner halo. We therefore colour the symbols of the bottom-left panel of Fig. 2 by the residuals about the running median of the binding energy of the halo as a function of halo mass. As motivated in Section 2.2, we use the intrinsic binding energy,  $E_{\text{DMO}}^{2500}$ , i.e. that recovered from each halo’s counterpart in the DMONLY simulation measured within  $r_{2500}$ . This eliminates the effects of dissipative physical processes, which could potentially mask or exaggerate any correlations induced by the intrinsic binding energy of the halo. The colouring reveals a striking negative correlation between the scatter in  $f_{\text{CGM}}$  and that of the binding energy of the inner halo.<sup>5</sup> The Spearman rank correlation coefficient is significant and negative for all  $10^{11.5} < M_{200} \lesssim 10^{13} M_{\odot}$ , and exhibits a broad, strong minimum (characterized by  $\rho < -0.5$  recovered at  $p < 0.01$ ) for  $M_{200} \lesssim 10^{12.5} M_{\odot}$ . The coefficient for haloes with  $M_{200} \simeq 10^{12.5} M_{\odot}$  is strong, with  $\rho' = -0.62$ . At fixed mass, more tightly bound haloes require more energy to unbind gas from the inner halo, so their central BHs must grow to be more massive, reaching a higher peak luminosity, and thus they ultimately eject a greater fraction of the halo gas beyond  $r_{200}$ .<sup>6</sup> More tightly bound haloes at fixed mass also tend to be those with a higher concentration and an earlier formation time (e.g. Navarro et al. 2004); indeed Booth & Schaye (2010, 2011) found that halo concentration correlates with  $M_{\text{BH}}$  at fixed  $M_{200}$ . We have therefore examined the relationships between the scatter in gas fractions at fixed halo mass and the scatter in the Navarro, Frenk & White (1997; ‘NFW’) concentration and the halo assembly lookback time (computed as per Qu et al. 2017) of each halo’s counterpart in the DMONLY-L100N1504 simulation, and recover negative correlations that are again significant, albeit slightly weaker than is the case for  $E_{\text{DMO}}^{2500}$ .

The binding energy of the halo in the DMONLY simulation is effectively encoded within the phase-space configuration of the initial conditions, and residuals of  $f_{\text{CGM}}$  correlate with similar strength, but over a wider range in halo mass, to those of  $E_{\text{DMO}}^{2500}$  than with those of  $M_{\text{BH}}$ . Scatter in  $E_{\text{DMO}}^{2500}$  at fixed  $M_{200}$  might reasonably then be considered as the fundamental cosmological origin of the scatter in  $f_{\text{CGM}}$  at fixed  $M_{200}$ . However, the influence of the binding energy is physically ‘transmitted’ to the gas fraction by ejective feedback. The necessity of this conduit can be demonstrated using the NOAGN-L050N0752 EAGLE simulation, in which AGN feedback is disabled. Examination of the relationship between  $\Delta f_{\text{CGM}}$  and  $\Delta \log_{10} E_{\text{DMO}}^{2500}$  in this simulation reveals a significantly weaker correlation than in the Reference simulation, which is driven largely by the lowest mass haloes in the sample. The origin of the correlation here differs with respect to the Reference simulation; haloes with greater binding energy (at fixed mass) exhibit lower gas fractions in the NOAGN simulation because the positive correlation of the  $\Delta f_{\text{gas}} - \Delta \log_{10} E_{\text{DMO}}^{2500}$  relation is much stronger than in the Reference simulation. The halo gas fraction is therefore depleted because of the condensation of gas into stars rather than its ejection by feedback.

The transmission of the influence of halo binding energy via AGN feedback can be more concisely demonstrated by examination of



**Figure 3.** Present-day halo baryon fractions,  $f_b$ , as a function of halo mass,  $M_{200}$ , in the Ref-L050N0752 (top) and NOAGN-L050N0752 (bottom) simulations. In each case the solid curve denotes the running median, and the symbols are coloured by the residuals about the running median of the  $E_{\text{DMO}}^{2500} - M_{200}$  relation,  $\Delta E_{\text{DMO}}^{2500}$ . Sub-panels show the running value of the Spearman rank correlation coefficient,  $\rho$ , for the relationships between these residuals and  $\Delta f_b$ , the residuals about the  $f_b - M_{200}$  relation. Shading denotes regions for which the correlation is recovered at low significance ( $p > 0.01$ ).

the halo baryon fraction,  $f_b \equiv [M_{\text{gas}}(r < r_{200}) + M_{\star}(r < r_{200})]/M_{200}$ . In Fig. 3, we show the present-day  $f_b - M_{200}$  relations for Ref-L050N0752 (top) and NOAGN-L050N0752 (bottom), and colour the symbols by  $\Delta \log_{10} E_{\text{DMO}}^{2500}$ . Using the halo baryon fraction rather than the halo gas fraction takes account of the aforementioned additional condensation of gas into stars in the NOAGN simulation. In the absence of AGN feedback, haloes with  $M_{200} \gtrsim 10^{12} M_{\odot}$  retain a significantly greater fraction of their baryons, as stellar feedback is unable to expel gas from massive haloes. In Ref-L050N0752, the  $\Delta f_b - \Delta \log_{10} E_{\text{DMO}}^{2500}$  correlation for haloes with  $M_{200} = 10^{11.5} - 10^{13} M_{\odot}$  is strong, significant, and negative for  $M_{200} > 10^{11.8} M_{\odot}$ , with  $\rho' = -0.66$ .<sup>7</sup> In NOAGN-L050N0752, the correlation is again mildly negative for  $M_{200} \lesssim 10^{12} M_{\odot}$ , but becomes moderately positive for more massive haloes, with  $\rho' = 0.43$ . Therefore, in the absence of AGN feedback, the intrinsic binding energy of the inner region of haloes has a markedly

<sup>5</sup>Using  $E_{\text{DMO}}^{500}$  or  $E_{\text{DMO}}^{200}$  instead yields similar results.

<sup>6</sup>This effect underpins the importance of using  $E_{\text{DMO}}^{2500}$  rather than  $E_{\text{Ref}}^{2500}$ . In general the dissipation of gas into stars results in  $E_{\text{Ref}}^{2500} > E_{\text{DMO}}^{2500}$ , but crucially the fractional difference in the two measures at fixed halo mass anticorrelates strongly with  $E_{\text{DMO}}^{2500}$ , because the formation of more massive BHs at the centres of haloes with greater  $E_{\text{DMO}}^{2500}$  enables them to eject a greater fraction of their gas. The use of  $E_{\text{Ref}}^{2500}$  therefore partially masks the underlying correlation between the inner binding energy and  $f_{\text{CGM}}$ .

<sup>7</sup>We compute  $\rho'$  for  $L = 50 \text{ cMpc}$  simulations within the larger window of 0.5 dex around  $M_{200} = 10^{12.5} M_{\odot}$ , to maintain adequate sampling.

different influence on the baryon fractions for  $M_{200} \gtrsim 10^{12} M_{\odot}$ . This notwithstanding, we reiterate that residuals about the  $f_{\text{CGM}}-M_{200}$  relation in the Reference model correlate strongly with those about the  $E_{\text{DMO}}^{2500}-M_{200}$  relation over a broader range in halo mass than with those about the  $M_{\text{BH}}-M_{200}$  relation, indicating that the binding energy may also influence halo gas fractions via other mechanisms, such as formation time.

It is interesting to note that the correlations between  $\Delta f_{\text{CGM}}$  and each of  $\Delta \log_{10} M_{\text{BH}}$  and  $\Delta \log_{10} E_{\text{DMO}}^{2500}$  change sign for  $M_{200} \gtrsim 10^{13} M_{\odot}$ , and become positive. The reversal of these correlations is likely a consequence of the declining efficiency of gas expulsion by AGN feedback in the most-massive haloes; as feedback becomes unable to eject gas from the assembling halo, a greater central binding energy only serves to inhibit gas expulsion.

Matthee & Schaye (2019) recently demonstrated that the star formation histories of EAGLE galaxies hosted by haloes with earlier formation times (at fixed halo mass) are systematically shifted to earlier cosmic epochs. A consequence of this effect is that these galaxies exhibit systematically lower present-day star formation rates (SFRs,  $\dot{M}_{\star}$ ). Since haloes with early formation times generally exhibit greater central binding energies, one might expect that scatter about the  $f_{\text{CGM}}-M_{200}$  relation will correlate with the SFR. Returning briefly to Fig. 2, we show in the bottom-right panel the  $f_{\text{CGM}}-M_{200}$  relation with symbols coloured by the residuals about the running median of the SFR as a function of halo mass,  $\Delta \log_{10} \dot{M}_{\star}$ , which indeed reveals a striking positive correlation between the residuals about the running medians of the  $f_{\text{CGM}}-M_{200}$  and  $\dot{M}_{\star}-M_{200}$  relations. The running coefficient of the Spearman rank correlation is positive for all halo masses, and for  $M_{200} \lesssim 10^{13} M_{\odot}$  it resembles the inverse of that of the  $\Delta f_{\text{CGM}}-\Delta \log_{10} M_{\text{BH}}$  correlation. The coefficient for haloes with  $M_{200} \simeq 10^{12.5} M_{\odot}$  is  $\rho' = 0.65$ , highlighting the strength of this correlation. More gas-rich haloes at fixed mass, besides exhibiting relatively undermassive BHs, therefore also exhibit an elevated SFR.

A close connection between the SFR of galaxies and their interstellar gas content is well established by observations (e.g. Kennicutt 1998), a finding whose reproduction and interpretation continues to attract considerable analytic and numerical effort (e.g. Thompson, Quataert & Murray 2005; Krumholz & Tan 2007; Semenov, Kravtsov & Gnedin 2016; Orr et al. 2018). To our knowledge, however, a correlation between the gas fractions of haloes (of similar mass) and the SFR of their central galaxies has not been demonstrated previously. The correlation of both the SFR and the BH mass with the halo gas fraction at fixed halo mass is helpful from the perspective of scrutinizing the predictions advanced here, as the SFR can be inferred from a diverse range of photometric and spectroscopic diagnostics; we turn to this scrutiny in the next section.

The correlation is also of intrinsic interest. Surveys of galaxies and associated absorption systems have revealed a positive correlation between the SFR of galaxies and the column density of the absorbers (e.g. Chen et al. 2010; Lan, Ménard & Zhu 2014; Rubin et al. 2018), and a popular interpretation of this correlation is that the absorption column densities are enhanced by outflows driven by stellar feedback. Analysis of EAGLE suggests that the correlation is not causal, but is rather a consequence of the negative correlation of both the halo gas fraction, and the SFR, with the halo binding energy at fixed halo mass. The first correlation is a consequence of more tightly bound haloes requiring more massive BHs to unbind gas from their inner regions, while the second is due to more tightly bound haloes collapsing earlier, shifting the growth of their central

galaxy and BH (and the associated expulsion of their halo gas) to earlier times. This interpretation is likely sensitive to the details of the phenomenological implementation of the relevant physical processes in the EAGLE model, and we defer further detailed exploration of this sequence of events to a follow-up study.

#### 4 TESTING VIA COMPLEMENTARY OBSERVABLES

The influence of central BH mass on the scatter of halo gas fractions at fixed halo mass is an unambiguous prediction of the EAGLE simulations. However, it is not one that is trivial to confront with observations, since one requires measurements of  $f_{\text{CGM}}$  and  $M_{200}$ , and dynamical measurements of BH masses, for a large sample of galaxies. We therefore briefly explore in this section whether it might be possible to test the predictions of Section 3 with complementary observational diagnostics provided by extant or forthcoming facilities.

We require diagnostics for which scatter about their running median as a function of halo mass correlates with scatter about the  $f_{\text{CGM}}-M_{200}$  relation. We first consider the diffuse, soft ( $0.5 \lesssim E_{\text{X}} \lesssim 2.0$  keV) X-ray luminosity of the hot ( $T \gtrsim 10^6$  K), collisionally ionized component of halo gas. Characterization of the properties of the gas in haloes less massive than those of galaxy groups ( $M_{500} \sim 10^{13} M_{\odot}$ ,  $kT \sim 1$  keV) remains challenging, with the extended hot CGM of only a handful of galaxies having been convincingly detected and characterized beyond the optical envelope of the galaxy (e.g. Dai et al. 2012; Bogdán et al. 2013b, 2017; Li et al. 2016, 2017), but forthcoming and proposed X-ray observatories such as *Athena* and *Lynx* promise to make such detections more commonplace. Moreover, stacking low spatial resolution *ROSAT* All-Sky Survey X-ray maps about the coordinates of local, optically selected galaxies has proven an effective means of characterizing the relationship between galaxies and their gas content (Anderson et al. 2015; Wang et al. 2016). We therefore compute the soft X-ray luminosity of each halo by coupling the physical properties of its constituent gas particles (i.e. those within  $r_{200}$ ) to the Astrophysical Plasma Emission Code (APEC; Smith et al. 2001), using the techniques described by McCarthy et al. (2017, see also Crain et al. 2010, 2013).

Stacking about the coordinates of optically selected galaxies has also recently been used to characterize the hot gas content of haloes via measurement of the thermal Sunyaev–Zel’dovich (tSZ) effect, the inverse Compton scattering of CMB photons by energetic electrons within the hot, ionized CGM (e.g. Planck Collaboration 2013; Greco et al. 2015). The tSZ ‘flux’ can be defined as the Compton- $y$  parameter integrated over the solid angle of the halo and is thus proportional to the total energy of the hot gas:

$$Y(< r_{200}) d_A(z)^2 = \frac{\sigma_T}{m_e c^2} \int_0^{r_{200}} 4\pi P_e(r, z) r^2 dr, \quad (1)$$

where  $d_A$  is the angular diameter distance to the halo,  $\sigma_T$  is the Thomson cross-section,  $m_e$  the electron rest mass, and  $P_e = n_e k_B T_e$  is the electron pressure with  $k_B$  being the Boltzmann constant. The flux therefore scales with the density of the hot gas, rather than the square of its density, as is the case for the collisional mechanisms that dominate the X-ray emissivity of diffuse plasmas. We compute  $Y$  by summing the contributions of the gas particles associated with each halo, as per McCarthy et al. (2017). Star-forming gas particles (i.e. those comprising the ISM) are assumed to be neutral and do not contribute to the flux.



The rows of Fig. 4 show the present-day  $L_X$ – $M_{200}$  (top) and  $Yd_A^2$ – $M_{200}$  (bottom) relations.<sup>8</sup> Histograms of  $L_X$  and  $Yd_A^2$  are shown to the right of the main panel. In each panel the solid black line denotes the LOWESS running median. In the left-hand column symbols are coloured by their residuals with respect to the running median of the  $f_{\text{CGM}}$ – $M_{200}$  relation. In both cases the colouring highlights that elevated values of the observable correspond to elevated values of  $f_{\text{CGM}}$ . The running values of the Spearman rank correlation coefficient demonstrate that scatter about both proxies correlates strongly, significantly and positively with that about  $f_{\text{CGM}}$  for all  $M_{200}$ . The X-ray luminosity is somewhat noisier than the tSZ flux, which is unsurprising since it is also sensitive to the metallicity of the halo (see e.g. Crain et al. 2013), and is more sensitive than the tSZ flux to the structure of the CGM.

The X-ray luminosity remains an attractive observable; however, owing in particular to the dynamic range it displays: for haloes with  $M_{200} \simeq 10^{12.5} M_\odot$  the 10th–90th percentile range spans 1.54 decades in  $L_X$ . Measurements of the X-ray luminosity therefore afford the opportunity to highlight readily the diversity of halo properties at fixed  $M_{200}$ . We demonstrate this quantitatively by showing, as grey curves, the median  $L_X$  of the subsets of haloes representing the upper and lower quartiles of the  $f_{\text{CGM}}$ – $M_{200}$  relation, in 10 bins of halo mass in the interval  $10^{11.5} < M_{200} < 10^{13} M_\odot$ . The separation of the subsets peaks at  $M_{200} = 10^{12.2} M_\odot$ , with the gas-rich subset of haloes exhibiting a median  $L_X$  that is 1.5 dex greater than that of the gas-poor subset.

To use  $L_X$  and  $Yd_A^2$  to test the influence of BH mass on halo gas fractions, they must respond to scatter about the  $M_{\text{BH}}$ – $M_{200}$  relation in a similar fashion to  $f_{\text{CGM}}$ . In the centre column of Fig. 4, the symbols are coloured by  $\Delta \log_{10} M_{\text{BH}}$ . Similar to the top-right panel of Fig. 2, residuals about the relations can be seen to correlate negatively with  $\Delta \log_{10} M_{\text{BH}}$ . The running values of the Spearman rank correlation coefficient, shown below the main panels, behave similarly and become more strongly negative from relatively low masses to a peak at  $M_{200} \simeq 10^{12.4} M_\odot$ . Therefore, the haloes of galaxies with central BHs that are more (less) massive than is typical for their mass are systematically ‘faint’ (bright) in both  $L_X$  and  $Yd_A^2$ . We again highlight the dynamic range of the X-ray luminosity and show as grey curves in the upper-centre panel the median  $L_X$  of the subsets representing the upper and lower quartiles of the  $M_{\text{BH}}$ – $M_{200}$  relation. The subsets are again most strongly separated at  $M_{200} = 10^{12.2} M_\odot$ , with the subset of haloes with undermassive BHs exhibiting a median  $L_X = 1.2 \times 10^{40} \text{ ergs}^{-1}$ , which is 1.4 dex greater than that of the overmassive BH subset,  $L_X = 4.5 \times 10^{38} \text{ ergs}^{-1}$ .

At present dynamical measurements of the masses of BHs are available for only  $\sim 10^2$  galaxies (e.g. Kormendy & Ho 2013; McConnell & Ma 2013), presenting an obstacle to observational scrutiny of the influence of BH mass on halo gas fractions. However, as noted in Section 3, the correlations of  $\Delta f_{\text{CGM}}$  with  $\Delta M_{\text{BH}}$  and  $\Delta \dot{M}_*$  are of similar strength (but opposite sign). Identification of this correlation in observations would therefore corroborate EAGLE’s predictions concerning the origin of scatter in the  $f_{\text{CGM}}$ – $M_{200}$  scaling

relation, hence it is important to establish how the proxies for  $f_{\text{CGM}}$  respond to scatter in SFR at fixed  $M_{200}$ .

In the right-hand column of Fig. 4 the symbols are coloured by  $\Delta \log_{10} \dot{M}_*$ . Encouragingly, we find that the residuals about the  $\dot{M}_*$ – $M_{200}$  relation correlate positively with the residuals about both the  $L_X$ – $M_{200}$  and  $Yd_A^2$ – $M_{200}$  relations, with the correlations being strong and significant for  $M_{200} \lesssim 10^{12.7} M_\odot$ . Therefore, the haloes of galaxies that exhibit an SFR that is high (low) for their halo mass are systematically ‘bright’ (faint) in both proxies. The grey curves of the upper-right panel show the median  $L_X$  of the subsets representing the upper and lower quartiles of the  $\dot{M}_*$ – $M_{200}$  relation. The subsets are again most strongly separated at  $M_{200} = 10^{12.2} M_\odot$ , for which the upper quartile ( $\dot{M}_* > 1.8 M_\odot \text{ yr}^{-1}$ ) exhibit a median  $L_X = 1.8 \times 10^{40} \text{ ergs}^{-1}$ , which is 1.7 dex (a factor of  $\simeq 50$ ) greater than that of the lower quartile ( $\dot{M}_* < 0.2 M_\odot \text{ yr}^{-1}$ ),  $L_X = 3.8 \times 10^{38} \text{ ergs}^{-1}$ .

## 5 SUMMARY AND DISCUSSION

We have examined the origin of scatter in the relationship between gas fraction,  $f_{\text{CGM}}$ , and mass,  $M_{200}$ , of the haloes with mass similar to those that host present-day  $\sim L^*$  central galaxies ( $10^{11.5} < M_{200} < 10^{13} M_\odot$ ) in the EAGLE simulations. We quantify the scatter by computing the difference between each halo’s gas fraction and the running median of the  $f_{\text{CGM}}$ – $M_{200}$  relation,  $\tilde{f}_{\text{CGM}}(M_{200})$ , constructed using the LOWESS locally weighted scatterplot smoothing method.

Our results are drawn primarily from the largest EAGLE simulation Ref-L100N1504, and its counterpart considering only collisionless gravitational dynamics, DMONLY-L100N1504. The parameters of the subgrid models governing feedback in the EAGLE Reference model were calibrated to ensure reproduction of key present-day properties of the galaxies, but the gaseous properties of galaxies and their haloes were not considered during the calibration and can be considered predictions of the simulations. We have also briefly studied the NOAGN-L050N0752 simulation in which AGN feedback is disabled, and its Reference model counterpart Ref-L050N0752.

Our findings can be summarized as follows:

(i) Scatter about the  $f_{\text{CGM}}$ – $M_{200}$  relation is not strongly correlated with residuals of the relationship between the stellar mass fraction of haloes and  $M_{200}$ . Low (high) halo gas fractions are therefore not generally a consequence of haloes having converted more (less) of their gas into stars throughout their assembly (Fig. 1).

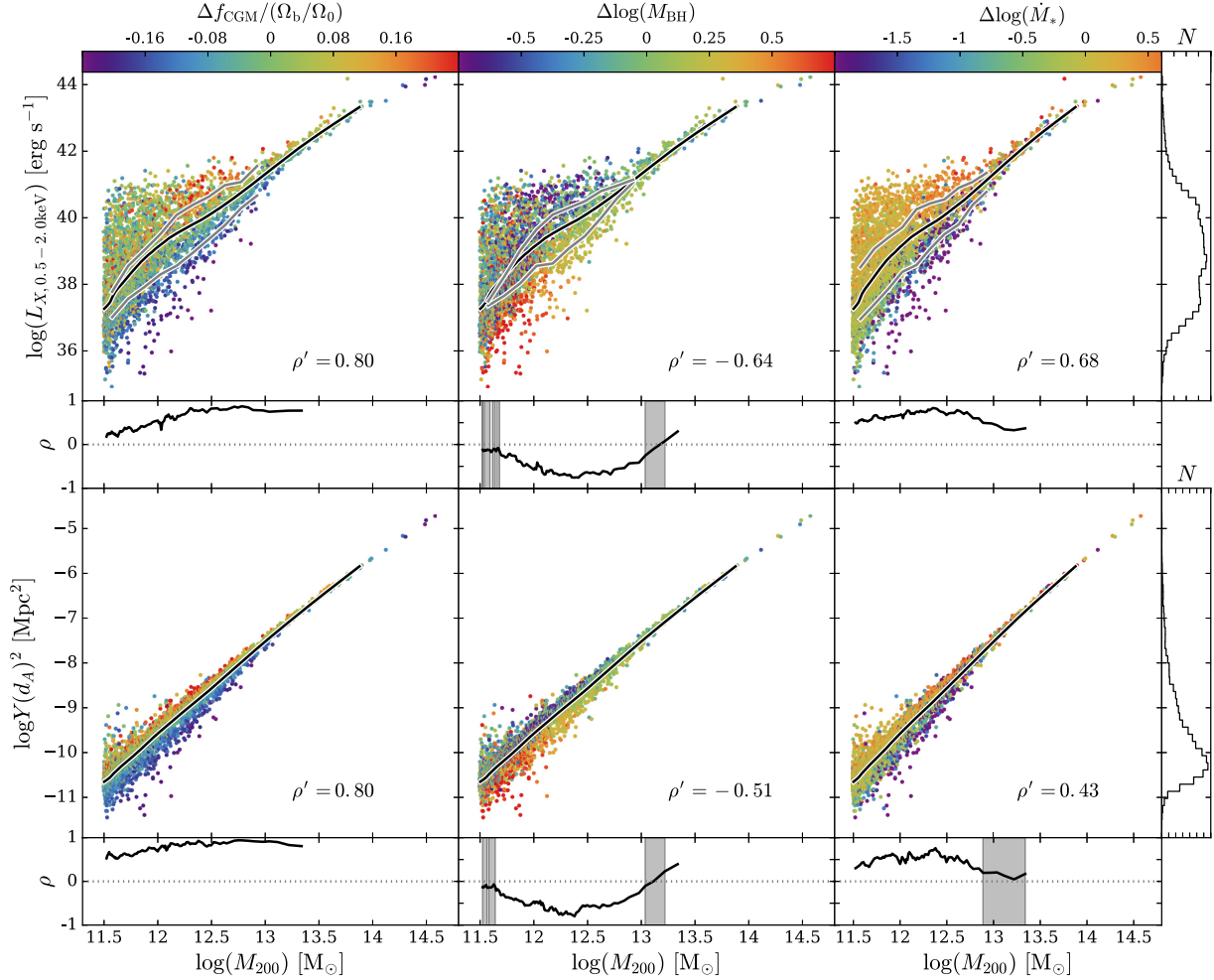
(ii) Similarly, the scatter is neither strongly nor significantly correlated with the residuals of the relationship between the gas accretion rate of central BHs (whether measured instantaneously or time-averaged over 100 Myr) and  $M_{200}$ . Low (high) halo gas fractions are therefore not a consequence of relatively strong (weak) ongoing AGN feedback (Fig. 2, top-left).

(iii) The scatter correlates strongly, significantly, and negatively with the residuals of the relationship between the present-day mass of central BHs and  $M_{200}$ . At  $M_{200} = 10^{12.5} M_\odot$  the Spearman rank correlation coefficient is  $\rho' = -0.75$ . At fixed  $M_{200}$  therefore, galaxies that host more-massive central BHs reside within relatively gas-poor haloes, and vice versa. This suggests that the main cause of scatter in  $f_{\text{CGM}}$  at fixed  $M_{200}$  is differences in the mass of halo gas expelled by AGN feedback throughout the assembly of the halo (Fig. 2, top-right).

(iv) A corollary of (ii) is the implication that the scatter about the  $f_{\text{CGM}}$ – $M_{200}$  relation might be driven by a more fundamental process that fosters scatter in the  $M_{\text{BH}}$ – $M_{200}$  relation. Booth & Schaye (2010,

<sup>8</sup>Since we are concerned with the response of  $L_X$  and  $Yd_A^2$  to deviations from the median scaling relations relating the properties of galaxies and their gaseous haloes to halo mass, precise correspondence between the properties of the simulated haloes and observational measurements is not crucial. Reasonable correspondence is however necessary to engender confidence in the realism of the simulations; we therefore compare the simulated scaling relations of  $L_X$  and  $Yd_A^2$  as a function of halo mass with observational measurements in Appendix A.





**Figure 4.** Present-day diffuse soft X-ray luminosity ( $L_X$ , *top*) and tSZ effect flux ( $Y d_A^2$ , *bottom*), as a function of halo mass,  $M_{200}$ . Histograms of  $L_X$  and  $Y d_A^2$  are shown to the right of the main panel. The solid curve denotes the running median of each quantity. In each column the symbols are coloured by the residuals about the running median of  $f_{\text{CGM}}$  (*left*),  $M_{\text{BH}}$  (*centre*), and  $\dot{M}_*$  (*right*) as a function of  $M_{200}$ . Beneath the main panels we show the running value of the Spearman rank correlation coefficient for the relationships between these residuals and  $\Delta f_{\text{CGM}}$ , the residuals about the  $f_{\text{CGM}}-M_{200}$  relation. Shading denotes regions for which the correlation is recovered at low significance ( $p > 0.01$ ), while  $\rho'$  denotes the value of the Spearman rank correlation coefficient computed for haloes with mass  $M_{200} \simeq 10^{12.5} M_\odot$ . The grey curves on the upper row show the median of the  $L_X-M_{200}$  relations of the halo sub-samples comprising the upper and lower quartiles of the diagnostic used for the symbol colouring.

2011) previously highlighted with cosmological simulations that this scatter is driven by differences in the binding energy of haloes at fixed mass. We find that scatter about the  $f_{\text{CGM}}-M_{200}$  relation indeed correlates strongly, significantly, and negatively with the residuals of the  $E_{\text{DMO}}^{2500}-M_{200}$  relation, where  $E_{\text{DMO}}^{2500}$  is the intrinsic binding energy of the halo, i.e. that which emerges in the absence of the dissipative physics of galaxy formation, measured within  $r_{2500}$ . This correlation is strong and significant over a broad range in  $M_{200}$ , and at  $M_{200} = 10^{12.5} M_\odot$  the correlation coefficient is  $\rho' = -0.62$ . (Fig. 2, bottom-left).

(v) Although reasonably interpreted as the fundamental origin of the scatter in  $f_{\text{CGM}}(M_{200})$ , the influence of the intrinsic binding energy of haloes is communicated via AGN-driven gas expulsion for  $M_{200} \gtrsim 10^{12} M_\odot$ . This is succinctly demonstrated by examination of the residuals about the  $f_b-M_{200}$  and  $E_{\text{DMO}}^{2500}-M_{200}$  relations (where  $f_b$  is the halo baryon fraction) in simulations with and without AGN feedback. In the Reference model simulation these residuals are strongly, significantly, and negatively correlated for all  $M_{200}$ , while in the NOAGN model the correlation is weaker for relatively

low-mass haloes ( $M_{200} \lesssim 10^{12} M_\odot$ ), and becomes positive for more massive haloes (Fig. 3).

(vi) The scatter in  $f_{\text{CGM}}(M_{200})$  correlates strongly, significantly, and positively with the residuals of the relationship between the present-day SFR of central galaxies and  $M_{200}$ . The correlation is similar to that with  $\Delta M_{\text{BH}}$ , but with opposite sign. At  $M_{200} = 10^{12.5} M_\odot$  the correlation coefficient is  $\rho' = 0.65$ . Haloes with high (low) gas fractions for their mass therefore typically host central galaxies with high (low) SFRs. (Fig. 2, bottom-right).

(vii) We consider the diffuse soft X-ray luminosity of the hot component of halo gas ( $L_X$ ) and the ‘flux’ of the thermal Sunyaev–Zel’dovich effect ( $Y d_A^2$ ) as proxies for  $f_{\text{CGM}}$ . Scatter about the relation of these observables with  $M_{200}$  correlates positively with residuals of the  $f_{\text{CGM}}-M_{200}$  relation, such that variations in the halo gas fraction at fixed halo mass are echoed by the two observables. Residuals about the running median of both proxies also correlate negatively with scatter about the median BH mass at fixed  $M_{200}$ . At  $M_{200} = 10^{12.5} M_\odot$  the associated correlation coefficients for  $L_X$  and  $Y d_A^2$  are  $\rho' = -0.62$  and  $\rho' = -0.51$ , respectively. This highlights

that they respond to variations in BH mass in a similar fashion to  $f_{\text{CGM}}$  (Fig. 4, left and centre columns).

(viii) Scatter about the  $L_X-M_{200}$  and  $Yd_A^2-M_{200}$  relations also correlates strongly and positively with scatter about the median SFR at fixed  $M_{200}$ . At  $M_{200} = 10^{12.5} M_\odot$  the associated correlation coefficients for  $L_X$  and  $Yd_A^2$  are  $\rho' = 0.68$  and  $\rho' = 0.43$ , respectively. These correlations afford a route to observational scrutiny of the predictions of the simulations advanced here, without the need to acquire dynamical BH mass measurements for a large sample of galaxies. The simulations indicate that, for galaxies hosted by haloes of  $M_{200} \simeq 10^{12.2} M_\odot$ , the median X-ray luminosity of those with  $\dot{M}_* > 1.8 M_\odot \text{ yr}^{-1}$  (the upper quartile in SFR) is a factor  $\simeq 50$  higher than for those comprising the lower quartile ( $\dot{M}_* < 0.2 M_\odot \text{ yr}^{-1}$ , Fig. 4, right-hand column).

The discovery of scaling relations connecting central BHs (with optical accretion discs of scale  $\sim 10^{-2}$  pc) with the properties of galaxies (on scales  $\sim 10^3$  pc, e.g. Magorrian et al. 1998; Kormendy & Ho 2013) has focused intense interest on the possibility of an intimate physical connection between the two. The release of rest mass energy from the accretion of gas on to BHs has long been advocated as a means to regulate cooling flows on to massive galaxies (e.g. Silk & Rees 1998) at the centres of groups and clusters (on scales  $\sim 10^5-10^6$  pc, e.g. Binney & Tabor 1995). Cosmological simulations still lack the physics and resolution required to capture the full complexity of the coupling between these phenomena across such a broad dynamic range, but our findings none the less indicate that central BHs can also have a significant influence on the structure and content of the CGM. It is an unambiguous prediction of the EAGLE simulations that scatter in the central BH mass, at fixed halo mass, markedly influences the gas fractions of the haloes that host present-day  $\sim L^*$  central galaxies.

Our findings also suggest that it is possible to corroborate or falsify EAGLE's predictions for the origin of scatter about the  $f_{\text{CGM}}-M_{200}$  relation, using extant or forthcoming observations. We posit that the locally brightest galaxy (LBG) sample from the New York University Value Added Galaxy Catalogue (Blanton et al. 2005), based on the seventh data release of the Sloan Digital Sky Survey (SDSS/DR7, Abazajian et al. 2009), is well-suited to this purpose. It has been used previously as the basis for stacked measurements of both diffuse X-ray luminosity (Anderson et al. 2015) from the *ROSAT* All-Sky Survey and the tSZ flux (Planck Collaboration 2013; Greco et al. 2015) from *Planck* maps. The acquisition of well-characterized rotation curves for this sample would further enable the identification of sub-samples with similar dynamical masses. We intend to pursue this line of enquiry in a forthcoming study; at present, X-ray luminosity remains the preferred proxy for  $f_{\text{CGM}}$ , as the  $\sim 10$  arcmin beam of *Planck*'s 100 GHz maps corresponds to a scale comparable to or larger than  $r_{200}$  for the majority of the LBG sample. However, characterization of the tSZ flux within lower mass haloes may soon be possible with the advent of next-generation, high-resolution, ground-based CMB experiments.

A further, complementary means of assessing variations in  $f_{\text{CGM}}$  is to search for variations in the column densities of absorption systems, revealed by the intersection of bright quasars with the CGM of a sample of nearby galaxies. Such an approach would be similar in spirit, but different in detail, to the COS-AGN survey (Berg et al. 2018). Extending the COS-Halos survey (Tumlinson et al. 2013), COS-AGN enabled a controlled comparison of the absorption systems associated with galaxies with and without AGN. Consistent with our finding that the present-day BH accretion rate has little impact on halo gas fractions at fixed  $M_{200}$ , Berg et al.

(2018) found no significant differences between these samples when examining the equivalent width distributions of absorption systems tracing the inner CGM. To probe the scenario advanced here, it is necessary instead to compare COS-Halos with a sample of galaxies with similar halo masses but diverse central BH masses (or a suitable proxy). Such a survey would also enable scrutiny of EAGLE's prediction that gas-rich (gas-poor) haloes exhibit higher (lower) SFRs than is typical for their halo mass, as a consequence of the negative correlation of scatter about both the  $f_{\text{CGM}}-M_{200}$  relation (due to the feedback history of the central BH) and the  $\dot{M}_*-M_{200}$  relation (due to the shift of the SF history to earlier times) with scatter about the  $E_{\text{DMO}}^{2500}-M_{200}$  relation. In a forthcoming study we will present expectations for the impact of BH mass and halo binding energy on the column densities and covering fractions of absorption species that are readily accessible in the local Universe, such as CIV.

## ACKNOWLEDGEMENTS

JJD acknowledges an STFC doctoral studentship. RAC is a Royal Society University Research Fellow. BDO is supported by NASA ATP grant NNX16AB31G. MS is supported by the Netherlands Organisation for Scientific Research (NWO) through VENI grant 639.041.749. SM acknowledges support from the Academy of Finland, grant number 314238. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 769130). The study used high performance computing facilities at Liverpool John Moores University, partly funded by the Royal Society and LJMU's Faculty of Engineering and Technology, and the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility ([www.dirac.ac.uk](http://www.dirac.ac.uk)). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grants ST/H008519/1 and ST/K00087X/1, STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure.

## REFERENCES

- Abazajian K. N. et al., 2009, *ApJS*, 182, 543
- Allen S. W., Schmidt R. W., Fabian A. C., 2002, *MNRAS*, 334, L11
- Anderson M. E., Churazov E., Bregman J. N., 2016, *MNRAS*, 455, 227
- Anderson M. E., Gaspari M., White S. D. M., Wang W., Dai X., 2015, *MNRAS*, 449, 3806
- Baldry I. K. et al., 2012, *MNRAS*, 421, 621
- Balogh M. L., Pearce F. R., Bower R. G., Kay S. T., 2001, *MNRAS*, 326, 1228
- Berg T. A. M., Ellison S. L., Tumlinson J., Oppenheimer B. D., Horton R., Bordoloi R., Schaye J., 2018, *MNRAS*, 478, 3890
- Binney J., Tabor G., 1995, *MNRAS*, 276, 663
- Bland-Hawthorn J., Gerhard O., 2016, *ARA&A*, 54, 529
- Blanton M. R. et al., 2005, *AJ*, 129, 2562
- Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, *MNRAS*, 456, 2361
- Bogdán Á., Bourdin H., Forman W. R., Kraft R. P., Vogelsberger M., Hernquist L., Springel V., 2017, *ApJ*, 850, 98
- Bogdán Á., Forman W. R., Kraft R. P., Jones C., 2013b, *ApJ*, 772, 98
- Bogdán Á. et al., 2013a, *ApJ*, 772, 97
- Booth C. M., Schaye J., 2009, *MNRAS*, 398, 53
- Booth C. M., Schaye J., 2010, *MNRAS*, 405, L1
- Booth C. M., Schaye J., 2011, *MNRAS*, 413, 1158
- Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
- Bower R. G., McCarthy I. G., Benson A. J., 2008, *MNRAS*, 390, 1399
- Bower R. G., Schaye J., Frenk C. S., Theuns T., Schaller M., Crain R. A., McAlpine S., 2017, *MNRAS*, 465, 32

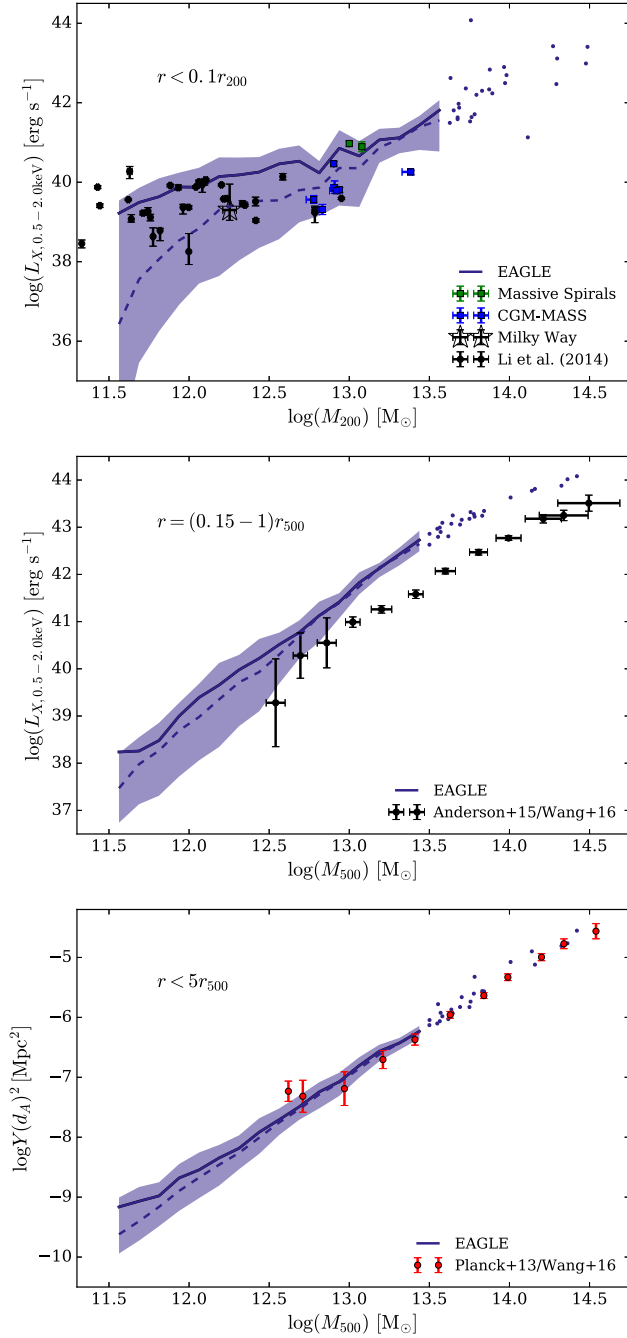
- Bregman J. N., 2007, *ARA&A*, 45, 221
- Chen H.-W., Wild V., Tinker J. L., Gauthier J.-R., Helsby J. E., Shtetman S. A., Thompson I. B., 2010, *ApJ*, 724, L176
- Cleveland W. S., 1979, *J. Am. Stat. Assoc.*, 74, 829
- Crain R. A., Eke V. R., Frenk C. S., Jenkins A., McCarthy I. G., Navarro J. F., Pearce F. R., 2007, *MNRAS*, 377, 41
- Crain R. A., McCarthy I. G., Frenk C. S., Theuns T., Schaye J., 2010, *MNRAS*, 407, 1403
- Crain R. A., McCarthy I. G., Schaye J., Theuns T., Frenk C. S., 2013, *MNRAS*, 432, 3005
- Crain R. A. et al., 2009, *MNRAS*, 399, 1773
- Crain R. A. et al., 2015, *MNRAS*, 450, 1937
- Crain R. A. et al., 2017, *MNRAS*, 464, 4204
- Croton D. J. et al., 2006, *MNRAS*, 365, 11
- Cullen L., Dehnen W., 2010, *MNRAS*, 408, 669
- Dai X., Anderson M. E., Bregman J. N., Miller J. M., 2012, *ApJ*, 755, 107
- Dalla Vecchia C., Schaye J., 2012, *MNRAS*, 426, 140
- Dolag K., Borgani S., Murante G., Springel V., 2009, *MNRAS*, 399, 497
- Durier F., Dalla Vecchia C., 2012, *MNRAS*, 419, 465
- Fabjan D., Borgani S., Tornatore L., Saro A., Murante G., Dolag K., 2010, *MNRAS*, 401, 1670
- Farahi A., Evrard A. E., McCarthy I., Barnes D. J., Kay S. T., 2018, *MNRAS*, 478, 2618
- Ford A. B. et al., 2016, *MNRAS*, 459, 1745
- Furlong M. et al., 2015, *MNRAS*, 450, 4486
- Furlong M. et al., 2017, *MNRAS*, 465, 722
- Gonzalez A. H., Sivanandam S., Zabludoff A. I., Zaritsky D., 2013, *ApJ*, 778, 14
- Greco J. P., Hill J. C., Spergel D. N., Battaglia N., 2015, *ApJ*, 808, 151
- Hopkins P. F., 2013, *MNRAS*, 428, 2840
- Hummels C. B., Bryan G. L., Smith B. D., Turk M. J., 2013, *MNRAS*, 430, 1548
- Kennicutt Jr. R. C., 1998, *ARA&A*, 36, 189
- Kormendy J., Ho L. C., 2013, *ARA&A*, 51, 511
- Krumholz M. R., Tan J. C., 2007, *ApJ*, 654, 304
- Lacey C., Cole S., 1994, *MNRAS*, 271, 676
- Lan T.-W., Ménard B., Zhu G., 2014, *ApJ*, 795, 31
- Li C., White S. D. M., 2009, *MNRAS*, 398, 2177
- Li J.-T., Bregman J. N., Wang Q. D., Crain R. A., Anderson M. E., 2016, *ApJ*, 830, 134
- Li J.-T., Bregman J. N., Wang Q. D., Crain R. A., Anderson M. E., Zhang S., 2017, *ApJS*, 233, 20
- Li J.-T., Crain R. A., Wang Q. D., 2014, *MNRAS*, 440, 859
- Lin Y.-T., Mohr J. J., Stanford S. A., 2004, *ApJ*, 610, 745
- Magorrian J. et al., 1998, *AJ*, 115, 2285
- Matthee J., Schaye J., 2019, *MNRAS*, 484, 915
- McAlpine S., Bower R. G., Harrison C. M., Crain R. A., Schaller M., Schaye J., Theuns T., 2017, *MNRAS*, 468, 3395
- McAlpine S., Bower R. G., Rosario D. J., Crain R. A., Schaller M., Schaye J., Theuns T., 2018, *MNRAS*, 481, 3118
- McAlpine S. et al., 2016, *Astron. Comput.*, 15, 72
- McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, *MNRAS*, 465, 2936
- McCarthy I. G., Schaye J., Bower R. G., Ponman T. J., Booth C. M., Dalla Vecchia C., Springel V., 2011, *MNRAS*, 412, 1965
- McCarthy I. G. et al., 2010, *MNRAS*, 406, 822
- McConnell N. J., Ma C.-P., 2013, *ApJ*, 764, 184
- Miller M. J., Bregman J. N., 2015, *ApJ*, 800, 14
- Moster B. P., Naab T., White S. D. M., 2013, *MNRAS*, 428, 3121
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493
- Navarro J. F. et al., 2004, *MNRAS*, 349, 1039
- Oppenheimer B. D., 2018, *MNRAS*, 480, 2963
- Oppenheimer B. D., Davé R., 2008, *MNRAS*, 387, 577
- Oppenheimer B. D., Davé R., Kereš D., Fardal M., Katz N., Kollmeier J. A., Weinberg D. H., 2010, *MNRAS*, 406, 2325
- Oppenheimer B. D. et al., 2016, *MNRAS*, 460, 2157
- Orr M. E. et al., 2018, *MNRAS*, 478, 3653
- Pillepich A. et al., 2018, *MNRAS*, 473, 4077
- Planck Collaboration, 2013, *A&A*, 557, A52
- Planck Collaboration, 2014, *A&A*, 571, A16
- Planck Collaboration, 2018, preprint (arXiv:1807.06209)
- Price D. J., 2010, *MNRAS*, 401, 1475
- Puchwein E., Sijacki D., Springel V., 2008, *ApJ*, 687, L53
- Puchwein E., Springel V., 2013, *MNRAS*, 428, 2966
- Qu Y. et al., 2017, *MNRAS*, 464, 1659
- Rahmati A., Schaye J., Bower R. G., Crain R. A., Furlong M., Schaller M., Theuns T., 2015, *MNRAS*, 452, 2034
- Rahmati A., Schaye J., Crain R. A., Oppenheimer B. D., Schaller M., Theuns T., 2016, *MNRAS*, 459, 310
- Rosas-Guevara Y. M. et al., 2015, *MNRAS*, 454, 1038
- Rubin K. H. R., Diamond-Stanic A. M., Coil A. L., Crighton N. H. M., Moustakas J., 2018, *ApJ*, 853, 95
- Schaller M. et al., 2015, *MNRAS*, 451, 1247
- Schaye J., Dalla Vecchia C., 2008, *MNRAS*, 383, 1210
- Schaye J. et al., 2010, *MNRAS*, 402, 1536
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Segers M. C., Schaye J., Bower R. G., Crain R. A., Schaller M., Theuns T., 2016, *MNRAS*, 461, L102
- Semenov V. A., Kravtsov A. V., Gnedin N. Y., 2016, *ApJ*, 826, 200
- Shull J. M., Smith B. D., Danforth C. W., 2012, *ApJ*, 759, 23
- Sijacki D., Springel V., 2006, *MNRAS*, 366, 397
- Silk J., Rees M. J., 1998, *A&A*, 331, L1
- Smith R. K., Brickhouse N. S., Liedahl D. A., Raymond J. C., 2001, *ApJ*, 556, L91
- Snowden S. L. et al., 1997, *ApJ*, 485, 125
- Springel V., 2005, *MNRAS*, 364, 1105
- Springel V., Di Matteo T., Hernquist L., 2005, *MNRAS*, 361, 776
- Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726
- Thompson T. A., Quataert E., Murray N., 2005, *ApJ*, 630, 167
- Trayford J. W. et al., 2015, *MNRAS*, 452, 2879
- Trayford J. W. et al., 2017, *MNRAS*, 470, 771
- Tumlinson J. et al., 2013, *ApJ*, 777, 59
- Turner M. L., Schaye J., Crain R. A., Rudie G., Steidel C. C., Strom A., Theuns T., 2017, *MNRAS*, 471, 690
- van de Voort F., Schaye J., 2012, *MNRAS*, 423, 2991
- Wang W., White S. D. M., Mandelbaum R., Henriques B., Anderson M. E., Han J., 2016, *MNRAS*, 456, 2301
- Werk J. K., Prochaska J. X., Thom C., Tumlinson J., Tripp T. M., O'Meara J. M., Peebles M. S., 2013, *ApJS*, 204, 17
- Wiersma R. P. C., Schaye J., Smith B. D., 2009a, *MNRAS*, 393, 99
- Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, *MNRAS*, 399, 574

## APPENDIX A: COMPARISON OF SIMULATED X-RAY LUMINOSITIES AND TSZ FLUXES WITH OBSERVATIONS

In this short Appendix we briefly compare the observable properties of haloes in the EAGLE simulation with observational measurements. In the panels of Fig. A1, we show the mean (solid lines) and median (dashed lines) values of scaling relations as a function of halo mass, with the 10th–90th percentile scatter shown via shading. In bins sampled by fewer than 10 galaxies, haloes are plotted individually. We remind the reader that, when comparing the simulations with measurements derived from stacked observational data, it is more appropriate to compare with the mean of the simulated sample rather than the median.

The upper panel of Fig. A1 focuses on the  $L_X$ – $M_{200}$  relation. The observational measurements are comprised of *XMM*–*Newton* and *Chandra* direct detections of diffuse emission from the environments of massive, isolated spiral galaxies, compiled and homogeneously reanalysed by Li et al. (2017). Specifically, these are observations of NGC 1961 (Bogdán et al. 2013a; Anderson,





**Figure A1.** Comparison of EAGLE’s CGM scaling relations with observational measurements. *Top panel:* the  $L_X$ – $M_{200}$  relation, where  $L_X$  is computed within a spherical aperture of  $r = 0.1r_{200}$ . Solid (dashed) lines indicating the binned mean (median) values, and shading indicates the 10th–90th percentile scatter. Galaxies in bins sampled by fewer than 10 galaxies are shown individually. Observational measurements are taken from the homogeneously reanalysed compilation of Li et al. (2017). *Centre panel:* the  $L_X$ – $M_{500}$  relation, where  $L_X$  is computed within a spherical annulus of  $0.15 < r/r_{200} < 1$ . Observational measurements are based on X-ray and weak lensing shear maps stacked about the coordinates of massive local galaxies, as per Wang et al. (2016). *Bottom panel:* the  $Yd_A^2$ – $M_{500}$  relation, where  $Y$  is computed in a spherical aperture of  $r = 5r_{500}$ . As above, observational measurements are based on the stacking analysis of Wang et al. (2016), here using *Planck* CMB maps.

Churazov & Bregman (2016) and NGC 6753 (Bogdán et al. 2013a), labelled here as ‘Massive Spirals’, the CGM-MASS sample of Li et al. (2017), a measurement of the Milky Way’s X-ray luminosity from Snowden et al. (1997) with uncertainty estimates from Miller & Bregman (2015), and a sample of inclined disc galaxies observed with *Chandra* and presented by Li, Crain & Wang (2014). Since direct detections of diffuse X-ray emission from galaxies are generally limited to  $r < 0.1r_{200}$ , the luminosities of EAGLE haloes in this panel were computed using a three-dimensional spherical aperture of this radius. The range of luminosities displayed by EAGLE galaxies is generally compatible with the scatter of the observational measurements, but we caution that direct detection observations are likely strongly biased towards the most X-ray-luminous galaxies at fixed mass.

In the central panel, we compare the X-ray luminosity of EAGLE galaxies to the  $L_X$ – $M_{500}$  scaling relation presented by Wang et al. (2016), who used weak lensing halo mass estimates to recalibrate the relation originally presented by Anderson et al. (2015), derived by stacking X-ray maps from the *ROSAT* All-Sky Survey about the coordinates of optically selected massive galaxies in the local Universe. The low spatial resolution of the maps precludes the excision of bright X-ray point sources, therefore we compare to the ‘CGM’ luminosities of Anderson et al. (2015), computed using an aperture of  $0.15 < r/r_{200} < 1$ . This comparison reveals that the extended gas haloes of very massive galaxies, and galaxy groups, in EAGLE are too X-ray-luminous. This shortcoming was previously highlighted by Schaye et al. (2015), who showed that the issue can be mitigated via the use of a higher temperature increment in EAGLE’s stochastic heating implementation of AGN feedback. We elected to present results here from the Ref-L100N1504 simulation rather than the AGNdT9-L050N0752, which uses this higher heating temperature, since the latter affords a factor of 8 poorer sampling of the galaxy population. Moreover, we focus here primarily on  $\sim L^*$  galaxies, for which the correspondence with the observed X-ray luminosity is reasonable for the Reference model.

Finally, we show in the bottom panel the stacked tSZ signal for the same sample used by Anderson et al. (2015) and Wang et al. (2016), as presented by Planck Collaboration (2013). Following McCarthy et al. (2017), we convert the observed values of  $Y(< r_{500})$  back into the measured flux  $Y(< 5r_{500})$ , and compare with tSZ flux of EAGLE galaxies computed within the same spherical aperture. The simulations reproduce the observed flux on these relatively large scales very well; a more stringent test of the simulation based on the tSZ flux awaits the availability of panoramic CMB maps with high spatial resolution.

We reiterate that the gaseous properties of galaxies and their haloes were not considered during the calibration of the parameters governing energetic feedback in EAGLE, and they may therefore be considered as predictions of the simulations. While it is possible to identify differences in detail, the observable properties of the extended gaseous environments of  $\sim L^*$  galaxies in EAGLE are in sufficiently good agreement with observational measurements to engender confidence in this aspect of the simulations.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.